The headline of a March 4, 2011 New York Times article seemed deliberately tailored to taunt lawyers’ worst fears about machine intelligence: “Armies of Expensive Lawyers, Replaced by Cheaper Software.” Beneath that headline was a photo of a lawyer sitting beside stacks of Bankers’ boxes. The caption read, “People get bored, people get headaches. Computers don’t.”

The accompanying story was only slightly less alarmist than its sensational headline. It warned: “Computers are getting better at mimicking human reasoning... and they are claiming work once done by people in high-paying professions.”

However, computers are unlikely to replace lawyers or any other legal professionals anytime in the foreseeable future. Even the smartest machines cannot emulate the judgment of a trained professional or function without human guidance. On the other hand, intelligent machines are increasingly becoming critical “members” of legal teams. They are not replacing professionals on the team; machines are making professionals more effective and efficient. Nowhere is this more evident than in e-discovery, where technology-assisted review (TAR) is enabling legal teams to handle cases otherwise too vast or costly to get through.

Having TAR or any other form of artificial intelligence (AI) on your team does not mean you have to take a computer to group lunches, but it does require you to re-examine your established tasks and workflows. What worked best in the past will change when AI joins your team, and some small accommodations for your electronic teammate can reap big rewards.

Best practices and potential pitfalls when integrating technology-assisted review and other advanced analytics into your existing processes.
A BRIEF BACKGROUND ON TAR
TAR is machine intelligence that uses software trained by human feedback to find relevant documents quickly and cost-efficiently. First introduced to the e-discovery market in 2009, TAR’s initial adoption was slow due to concerns about its defensibility. With its continued evolution and uniform acceptance by the courts, TAR is now part of the e-discovery mainstream.

The first TAR processes (TAR 1.0) were designed to dramatically reduce review time and costs with the promise of transforming the economics of e-discovery, particularly in big data cases. The initial technology minimized the volume of data and intelligently analyzed content, reducing the need for manual review and providing a more accurate picture of the data. TAR 1.0 systems were limited to providing a static process. Because sampling for both training and measuring performance happened only once in the beginning of the review process, there was no easy way to accommodate rolling data uploads or to use reviewers’ ongoing judgments to improve the accuracy of the algorithm.

Next-generation versions of TAR (TAR 2.0) take cost savings and intelligent review further. TAR 2.0 allows for continuous learning throughout the review process. Review team judgments can be continuously fed back into the system, enabling the system to get smarter about the document population and improve accuracy. TAR 2.0 also allows additional documents to be added at any time. Continuously updated ranking produces better results, meaning fewer documents to review, fewer attorney reviewers, decreased review time and increased cost savings.

BEST PRACTICES FOR WORKING WITH TAR
While TAR 2.0 is exceptionally good at some tasks, it is not great at others. For legal professionals, part of the challenge is knowing which tasks to transfer to this technological team member while still maintaining control over the overall process.

Pre-TAR workflows assume human decision-making at every stage, which means these workflows require thoughtful adjustments to make best use of newer technologies. As a starting point, it is helpful to separate tasks into those that humans do best (such as reading for comprehension or making relevance judgments) and those that machines do best (such as recognizing patterns across large volumes of data).

TAR can assist legal professionals in decision-making about data, but legal professionals must still make final decisions. Implementing best practices can help you achieve optimal output from TAR and make optimal use of human judgments. Some examples of best practices include:

- Intelligent reuse of expensive human judgments
- Storing only one measure per data field (e.g., whether a document is responsive, as opposed to whether it is subject to production)
- Coding documents on a document level, not a family level (the machine can’t tell whether you intended “responsive” to mean “responsive” or “subject to production”)

TAR USE BY TASKS
Let’s look at the kinds of tasks we face in e-discovery. Broadly speaking, document review tasks fall into three categories:

- **Classification**: The most common form of document review in which documents are sorted into buckets, such as responsive or non-responsive
- **Protection**: A higher level of review in which the purpose is to protect certain types of information from disclosure (the most common example is privilege review)
- **Knowledge-Generation**: Learning what stories the documents tell and discovering information useful to our case

Although TAR is helpful for all three, each has different metrics for success. Those metrics have important implications for designing your workflows and integrating TAR.

Recall and precision are two crucial metrics for measuring the effectiveness and defensibility of TAR processes. Recall is a measure of completeness, the percentage of relevant documents retrieved. Precision measures purity, the percentage of retrieved documents that are relevant.

The higher the percentage of each, the better you have done. If you achieve 100 percent recall, you have retrieved all the relevant documents. If all the documents you retrieve are relevant, you have achieved 100 percent precision. But recall and precision are not friends; typically, a technique that increases one will decrease the other. The three categories
documents. TAR can be of critical help in prioritizing the document population by issue and concentrating the most interesting documents at the top of the list so attorneys can quickly learn what they need to litigate the case.

One problem is that TAR algorithms rank documents according to their likelihood of getting a thumbs-up or thumbs-down from a human reviewer. They do not rank documents based on their degree of interest. Some documents could be easy to predict as responsive, but not very interesting. Other documents could be extremely interesting, but harder to predict because they are so unusual.

In practice, however, the more interesting documents cluster near the top of the ranking. Interesting documents sort higher this way because they contain stronger terms and concepts and more of them. TAR’s ability to concentrate the interesting documents near the top of a ranked list makes it a useful addition to knowledge-generation workflows.

MACHINES AND HUMANS IN HARMONY

This framework can help you think about, develop and evaluate different discovery workflows when machine intelligence joins your team. The critical factor in your success will be designing workflows that most effectively use all the tools and resources at your disposal. TAR is a powerful addition to your team — welcome the machines aboard!

Recall and precision are two crucial metrics for measuring the effectiveness and defensibility of TAR processes.

of document review tasks have different recall and precision targets, so choose and tune your workflows for each to maximize effectiveness and minimize cost and risk.

CLASSIFICATION TASKS: When using TAR for document production, classify documents so you can do different things — such as review, discard or produce — with subpopulations. The goal of using TAR is to get better results, not perfect results. You want to achieve reasonably high percentages of recall and precision, but at levels of cost and effort that are proportionate to the case. A goal of 80 percent recall (a common TAR target) could be reasonable when reviewing for responsive documents.

Precision must also be reasonable, but requesting parties are usually more interested in making sure they get as many responsive documents as possible. Recall usually takes precedence.

PROTECTION TASKS: When your task is to protect certain types of confidential information, you need to achieve 100 percent recall — nothing can fall through the cracks. This is problematic in practice.

To approximate perfection, you need to adjust the workflow to use every tool in your toolkit — not just TAR, but also keyword searching and human review — to identify the documents that must be protected.

The reason for this is simple: Different review methods make different mistakes. Human reviewers tend to make random mistakes. TAR systems typically make systematic errors, getting entire classifications of documents right or wrong. By combining different methods into the workflow, each serves as a check against the others.

Because TAR does not make the same class of errors as search terms and human review, it makes a valuable addition to privilege and other data protection workflows, provided the technology can deal with low prevalence and be efficiently deployed.

Precision is somewhat less important when your task is to protect documents. However, protection workflows include much human review, so including unnecessary junk quickly gets expensive. You want to achieve a fairly high level of precision, but recall is still the metric to focus on.

KNOWLEDGE-GENERATION TASKS: The final task is where the name “discovery” originated. What stories do these documents tell? What can we learn from them? For knowledge-generation, we do not particularly care about recall. We do not want all the documents about a topic, just the best ones — the ones that will end up in front of deponents or used at trial.

Precision and relevance are therefore the most important metrics. You do not want to waste your time going through junk, duplicative or less relevant